REPORT OF THE HIGH PERFORMANCE COMPUTER TASK FORCE

MARCH 6, 1992

Alberta
PREMIER'S COUNCIL ON SCIENCE AND TECHNOLOGY

REPORT OF THE HIGH PERFORMANCE COMPUTER TASK FORCE

MARCH 6, 1992

INTRODUCTION

This Report was prepared in the summer and fall of 1991 by a Task Force of the Premier's Council on Science and Technology, under the chairmanship of Dr. Marshall M. Williams (see Attachment One for names of Task Force members).

The Report consists of two parts - the Report itself constitutes Part One, while Part Two provides additional background information.

Part One consists of some definitions of high performance computing (page 2), a discussion of two basic considerations (page 3), a summary background presentation (pages 4 to 12), and conclusions (pages 13 to 15), followed by two appendices (pages 16 to 18).

DEFINITION OF HIGH PERFORMANCE COMPUTING

High performance computing involves far more than a supercomputer. It is a culture and an infrastructure consisting of five essential parts, of equal priority.

1. Hardware Systems - these extend beyond supercomputers to encompass peripheral gear and a hierarchy of related computing equipment (which may include for example specialized engines such as graphics devices, database servers, communications gear, and lower level computers to feed the specialized units and communications equipment).

2. Software - the more advanced or more comprehensive the system, the more complex the software required to operate the system and to exploit the power within it. The term "software" includes both operating systems and applications software.

3. Educated Users - a large cohort of the system user population must be familiar with the utilization of supercomputing, its applications, and proper use of resources to exploit the system. Ideally, this understanding extends to an even larger portion of the population who may not use or have physical contact with high performance computing but who appreciate the contribution it makes to the economic well being of the community.

4. Network - high performance computing requires the broad interaction of all the above parts; users, hardware, and software over network links capable of man-machine, machine-machine communication and of transporting large amounts of data at near instantaneous speeds.

5. Support Structure - the human and other support systems and the people who make the system operate so that it is friendly to users. They are involved in managing the hardware and network, in assisting users to better access the system, in training, in maintenance, etc.

High performance computing is an overall environment and strategy, not any specific class of machine. The process involves utilizing a range of computers, including a supercomputer, each performing tasks for which it is best suited.

Alberta currently has a partial high performance computing base. It is the purpose of this report to recommend how a complete high performance computing environment can be achieved from this base. We do have good workstation, mini-supercomputer and mainframe capabilities and a reasonable base of qualified personnel. However, we lack a supercomputer and certain aspects of the high-speed communication network essential to fully implement a high performance computing environment.

TWO BASIC CONSIDERATIONS

THE CONCEPT OF INFRASTRUCTURE

Each century adds unique infrastructure to that which has gone before it.

Thus, in the 19th century, governments owned and/or subsidized the operation of canals, railways, telegraph lines, harbours, water and sewer lines.

In the 20th century, this ownership/subsidization was extended to telephones, electric power, pipelines, highways, airports, etc.

In the final decade of the 20th century, and into the 21st century, high speed communications networks coupled with high performance computing must be added to the basic infrastructure of those jurisdictions which expect to be globally competitive in the information age.

The major difference in the newest transportation/communication infrastructure, compared to the existing infrastructure, is the very rapid rate of development and improvement of many of the components, measured in a very few years, as compared to decades for the more traditional infrastructure.

THE RAPID DEVELOPMENT OF HIGH PERFORMANCE COMPUTING POWER

It is important to recognize that most of today's routine industrial computing operations were the high performance computing developments of but a few years ago.

Airline reservation systems, project scheduling, CAD/CAM design, seismic data processing, satellite image processing, and computer modelling of all kinds, including petroleum reservoir simulations, were only yesterday beyond the reach of available software and hardware. Today, many businesses cannot function without them.

High performance computing developments today pave the way for tomorrow's routine commercial and research applications.

BACKGROUND ON HIGH PERFORMANCE COMPUTING

1. THE NICHE ROLE OF HIGH PERFORMANCE COMPUTING IN THE DOMAIN OF ELECTRONIC COMPUTING

- the term "high performance computing" actually describes an overall environment wherein various types of computers, powerful in their own specific field, function in a cooperative manner, to achieve optimum results. Supercomputers are at the leading edge of the computing domain - they pace the development of the computational disciplines - and they are continually being further developed to reach new levels of performance (see Figure One)

- the pace of development is such that today's supercomputer will be "tomorrow's" high end mainframe and "next week's" workstation

- the cost of computers does not change significantly over time in constant dollars - however, the cost per operation performed has declined dramatically with time (see Figure Two) - for example, today's supercomputer is capable of one billion operations per second, but will probably be replaced by supercomputers by the year 2000 (if not before) capable of one trillion operations per second

- the demand by users, for increased computational speed grows as fast as computer capacity grows

- as long as computers exist, there will always be a class of machines known as supercomputers, and there will always be companies filling this demand - the current leaders in this field are Cray, Fujitsu, Hitachi, IBM and NEC

- there are certain types of problems which not even a supercomputer of the year 2000 will likely be able to solve quickly, and with the desired precision - those of greatest interest in Alberta are:

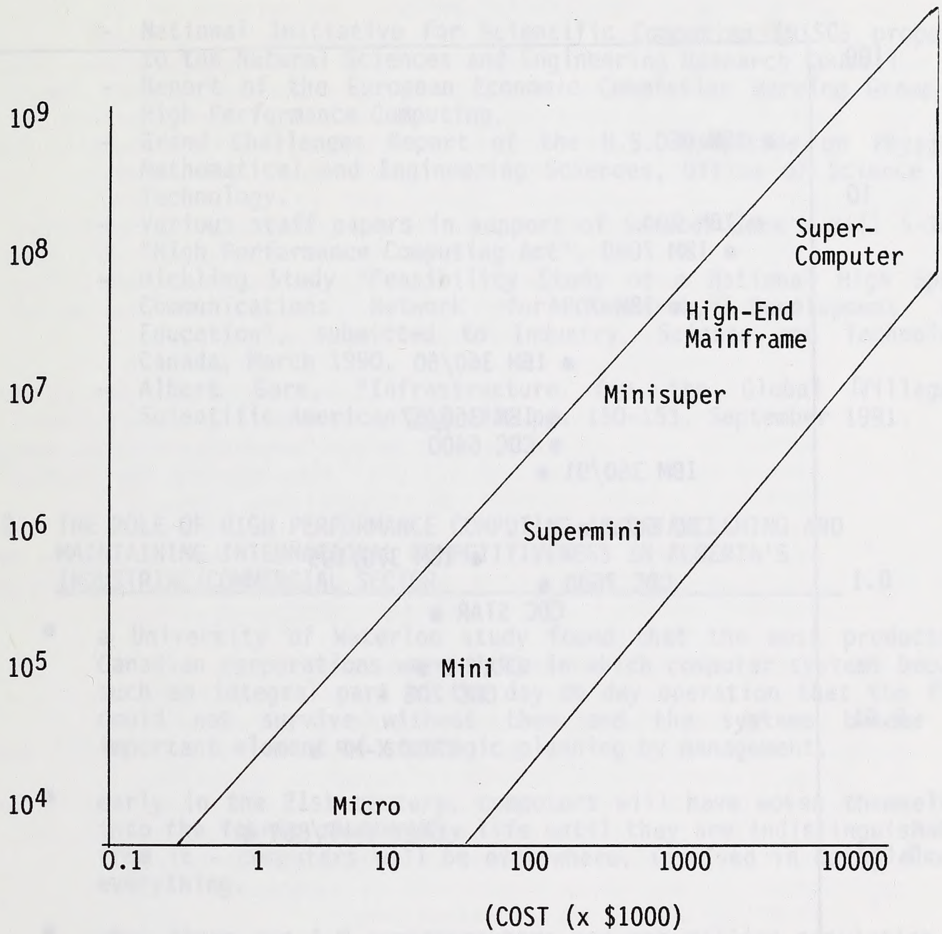
  - signal processing of seismic data
  - petroleum reservoir modelling and simulation
  - long-range climate modelling
  - drug design
  - structure of molecules
  - cosmological problems; etc.

  Thus, supercomputers do not fill a "niche" role - rather they deal with a whole spectrum of complex problems in many different industrial and academic areas of interest.

- new technology in the form of "massively parallel" computer architecture will be introduced to meet the demands of high performance computer users by the 21st century - this development will not obviate the need for existing supercomputer architectures but will supplement them and be integrated into high performance computer networks

Figure One

PERFORMANCE AND COST OF AVAILABLE TYPES OF COMPUTERS

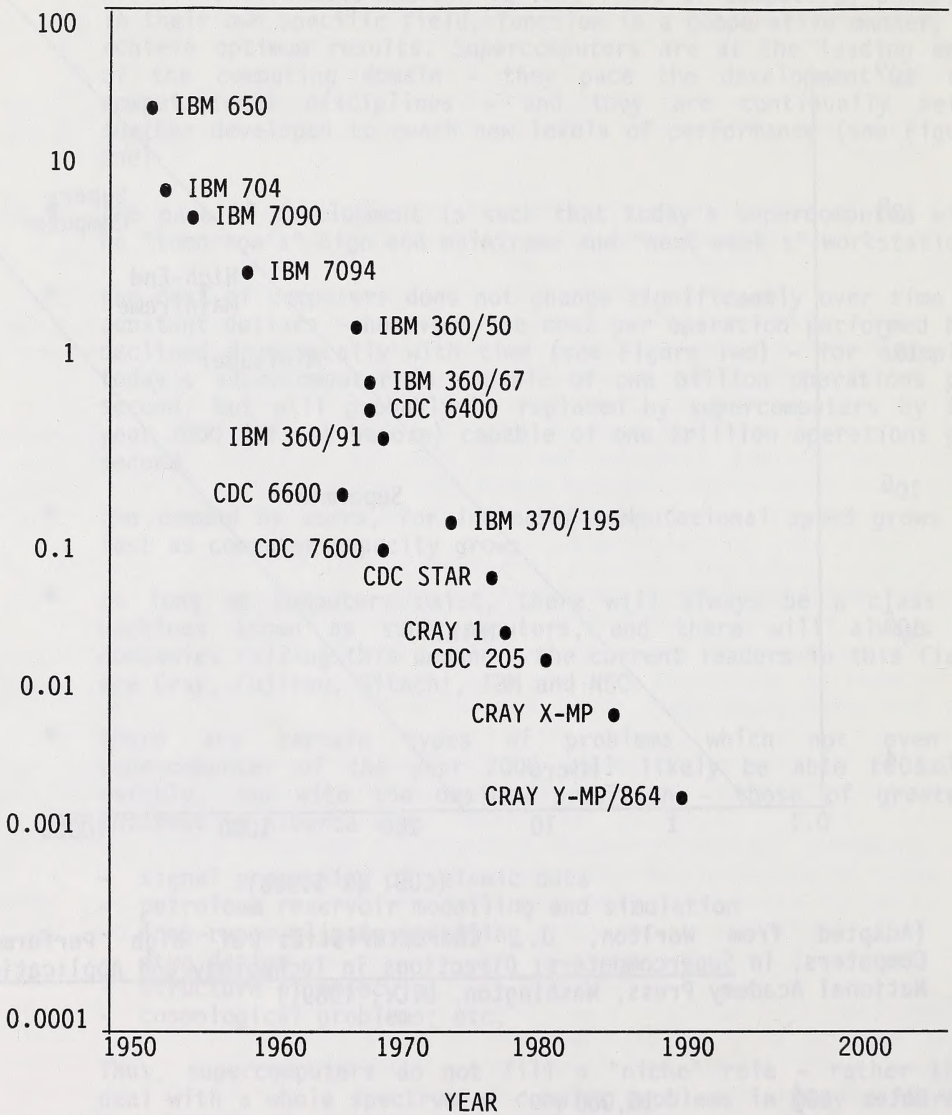


(Adapted from Worlton, J., Characteristics of High Performance Computers, in Supercomputers: Directions in Technology and Applications, National Academy Press, Washington, D.C., 1989.)

Note: 10^4 = 10,000
10^5 - 100,000
10^6 = 1,000,000
10^7 = 10,000,000
10^8 = 100,000,000
10^9 = 1,000,000,000

Figure Two

DECLINING COSTS OF COMPUTING



Note: Adapted from "Scientific American", September 1991, p. 88.

Note: The machine at the Calgary Supercomputer Centre is the CDC 205.

- the necessity of high performance computing in a growing and healthy economy has been stated vigorously in the presentations made to the Task Force and in such articles as the:

  - National Initiative for Scientific Computing (NISC) proposed to the Natural Sciences and Engineering Research Council
  - Report of the European Economic Commission Working Group on High Performance Computing.
  - Grand Challenges Report of the U.S. Committee on Physical Mathematical and Engineering Sciences, Office of Science and Technology.
  - Various staff papers in support of Senator Gore's Bill S-1076 "High Performance Computing Act".
  - Hickling Study "Feasibility Study of a National High Speed Communications Network for Research, Development and Education", submitted to Industry, Science and Technology Canada, March 1990.
  - Albert Gore, "Infrastructure for the Global Village", Scientific American 265, #3, pp. 150-153, September 1991.

2. THE ROLE OF HIGH PERFORMANCE COMPUTING IN ESTABLISHING AND MAINTAINING INTERNATIONAL COMPETITIVENESS IN ALBERTA'S INDUSTRIAL/COMMERCIAL SECTOR

- a University of Waterloo study found that the most productive Canadian corporations were those in which computer systems became such an integral part of the day to day operation that the firm could not survive without them and the systems became an important element of strategic planning by management.

- early in the 21st century, computers will have woven themselves into the fabric of everyday life until they are indistinguishable from it - computers will be everywhere, involved in doing almost everything.

- when there are 1.2 supercomputers per one million population in Japan, and 0.6 supercomputers per one million population in the U.S.A., versus only 0.2 supercomputers per one million population in Canada, it is obvious that we must have access to, and be capable of using supercomputers, if Alberta is to be industrially competitive in a global sense in this decade and beyond - many of the foreign companies with which Alberta companies compete use supercomputers.

- Alberta is the only province in Canada in which a commercially owned and operated supercomputer exists, i.e., Shell Canada Resources, Calgary - there are also currently a number of other high performance computers in the Province.

- some oil companies located in Alberta (Amoco, Chevron, Esso) also have network access to their parent company's supercomputers in the U.S.A. - but some other companies undertake the vast majority of their research work outside Alberta because they do not have access to a supercomputer here.

- in Japan, over half the country's supercomputers are owned by industrial firms, and universities own slightly over one-quarter of them - Fujitsu alone has 68 installed supercomputers in Japan with over 100 on order - against competition like this, Alberta companies clearly need ready access to a supercomputer.

- smaller oil companies, to become and remain competitive need access to a supercomputer for their seismic and oil reservoir work, because the major companies are already utilizing this technology to reduce their costs; if the smaller ones do not also do so, they will have difficulties in surviving.

- while drug design in Alberta is currently done at the University of Alberta, much of it is under contract to international drug companies - it is likely that before too long, Alberta companies like Biomira, Synphar and SPI will also require the capabilities that only a supercomputer can supply.

- supercomputers are very useful in silicon chip circuitry design (such as is done by LSI Logic Canada) and in the design of advanced materials, (e.g., Westaim, EDO).

- other potential industrial usages are:

  (i) studying flow and behavior in pipelines;
  (ii) studying turbulence and behavior in reactors in oil and petrochemical industry plants;
  (iii) use in complex graphics, animation and visualization (often the only practical way to interpret seismic data), etc.
  (iv) in telecommunications infrastructures, etc.

- potential public sector usages are:

  (i) developing models of the Alberta and Canadian economies for economic forecasting;
  (ii) studying effects of environmental change and weather forecasting;
  (iii) modelling health trends;
  (iv) applications in education;
  (v) applications in agricultural, e.g., effect of new pesticides, global warming, etc.

- supercomputers provide industry and academia with a competitive edge in leading domains

- people using supercomputers are today's leaders, who are shaping our tomorrow

3. THE ROLE OF HIGH PERFORMANCE COMPUTING, ESPECIALLY IN RESEARCH AND DEVELOPMENT, IN ALBERTA'S ACADEMIC SECTOR

- the proportion of university researchers who use supercomputers is relatively low, yet these researchers tend to be among the most prestigious, those with many outside awards, those with outstanding world-wide recognition, and those with a disproportionately large number of graduate students - their work is representative of the most advanced ideas in Canada - this is our "national distinctive competence" and is what sets us apart from other jurisdictions - it is the "food" for developing new technologies for the future, and hence one of the sources of provincial and national income in the 21st century.

- they are working in fields such as:

  - signal processing of seismic data
  - drug design and drug action on human tissues
  - global environmental effects
  - structure of chemical molecules
  - interaction of sub-atomic particles
  - cosmology
  - economic modelling
  - voice recognition, etc.

- while supercomputers themselves are expensive, they have two significant capabilities:

  - they make possible, solutions to problems which can't otherwise be done (or can't effectively be done) in a number of areas (see bullet immediately above); and,
  - they lower the cost of finding solutions where funding solutions through experiments are very costly or impossible, e.g., in the failure of structures such as Arctic ice platforms, off-shore drilling rigs, aircraft, buildings in earthquake zones, the behavior of materials inside reactors in the petrochemical and oil industries, etc.

4. THE RELATIVE EFFECTIVENESS OF LARGE COMPUTERS VERSUS SMALLER, MORE SPECIALIZED ONES, IN MEETING ALBERTA'S NEEDS

- the question has been asked: Why not buy one or more mini-supercomputer machines at $2 million or so each, versus one supercomputer costing $10-15 million?

- the answer is that mini-supercomputers, workstations, etc., address some parts of problems and are needed as part of a total high performance computing facility, particularly for the visualization of complex results - but they are ancillary to the heavy duty computing performed by a supercomputer - such devices are also used to prepare the programs and data that are executed on the largest processors - this is because a human being interfacing with a supercomputer cannot work nearly as quickly as the supercomputer can process data - in order to make a supercomputer run as efficiently as possible, data is fed into a small computer, which then feeds data rapidly into a supercomputer and uses the latter's capabilities more effectively than direct human intervention can.

- Our investigation clearly established that the issue is not really a "relative" one. Both large computers and specialized ones have an important role. However we simply have to conclude that specialized parallel computers cannot take over the role of large computers at this time. We will not know how effective specialized parallel computers will be for three to five years. In the meantime, Alberta should not wait as we will lose ground in our economic development and in particular the development of essential professional talent. Even worse we could lose some of the talent we now have.

- the situation is analogous to two contractors - one with 50 one-ton trucks, and the other with one 50-ton truck - under some circumstances the one-ton trucks are more effective - under others, both contractors may be competitive - but if a 50-ton rock has to be moved, the contractor with the one-ton trucks is not in the game - or, as another example, a model of a long-distance pipeline may take two to three hours to calculate on a supercomputer, but two to three weeks on a mini-supercomputer.

- with larger problems, and the greater the scope of a solution to a problem, the less likely one is to be surprised by the results, and the greater the benefit from the wider optimization that is possible.

- several large-scale problems are in areas of economic importance to Alberta which is why we have the only privately owned supercomputer in Canada, and a high concentration of high performance computers.

- nevertheless, even in the year 2000, when supercomputers capable of over a trillion operations per second are expected to be available, seismic data reduction problems could still make them choke.

5. HIGH PERFORMANCE COMPUTING'S ROLE IN RETAINING WORLD-CLASS SCIENTISTS AND ENGINEERS

- a supercomputer facility is getting to be a desired part of research facilities - its presence is an indication of a leading institution and will serve to retain and attract world-class researchers - its absence is beginning to be a negative reflection on the competitiveness of institutions.
- high performance computing facilities are not as rare as supercomputer facilities - they are expected to be available in every significant research facility that requires data manipulation - consequently, their existence will not serve to retain or attract world-class researchers.
- Canada and Alberta have a high proportion of talented software developers, who need access to more powerful hardware than they now have.

6. SOME QUOTES ON HIGH PERFORMANCE COMPUTING

EUROPEAN ECONOMIC COMMISSION

"Scientific and societal progress, industrial competitiveness, the understanding and control of environmental factors necessary to human well-being will be governed by the availability of adequate computing power."

> Report of the European Economic Commission Working Group on High Performance Computing

U.S. SENATOR ALBERT GORE - Note: Senator Gore's Bill passed both the Senate and the House of Representatives, unanimously, in September 1991

> "If we're going to build an economy strong enough to grow in an information age dominated by high technology and tough foreign competitors, these are investments (i.e., in a national information super-highway linking powerful supercomputers through a high speed network and supporting critical research in high performance computing) we must make now. In every sector of our economy - from finance to farming - these technologies could make a difference."

> Adapted from Senator Gore's introduction of the High Performance Act to the U.S. Senate January 24, 1991.

"High performance computing represents a future that's already arriving. The U.S. lead in high tech industries is being challenged and, because of indifference, or inaction, could be overtaken, leaving us to pay a high price in lost jobs and lost opportunities in every sector of our society."

Sen. Gore, Thursday, January 24, 1991 on introduction of the High Performance Computing Act (passed both Houses, September 1991).

"New technologies that enhance the ability to create and understand information have always led to dramatic changes in civilization. The printing press unleashed the forces that led to the birth of the modern nation state. It made possible the widespread distribution of civic knowledge that enabled the average citizen to affect political decisions.

Now come distributed networks connecting a myriad of computers, ranging from megaflop machines and work stations to desktop and personal laptop units. All are becoming less expensive in each successive generation that emerges. There is no longer any doubt that such machines will reshape human civilization even more quickly and more thoroughly than did the printing press."

Sen. Gore "Infrastructure for the Global Village", Scientific American, September 1991, pp. 150-153.

"The nation which most completely assimilates high performance computing into its economy will very likely emerge as the dominant intellectual, economic and technological force in the next century".

Sen. Gore on the introduction of Bill S-1067 in the U.S. Senate, May 18, 1989.

CONCLUSIONS

- Alberta will not be able to continue to generate the wealth to maintain itself at its present standards through its natural resources as it has done in the past.

- Therefore, an alternative means to create wealth must be developed. One way to do this is through the development and application of technology based on the ingenuity of our people.

- Those who accomplish this will do so through the application of knowledge, superior reasoning and strategic action. Such people represent our distinctive, uniquely Canadian, national competence.

- Computer applications enhance our ability to know, reason and act, and this enhancement appears to be proportional to the size and power of available computer systems.

- Computer and computer applications also affect the way business and research are done, and how we communicate, not only amongst ourselves but also with our global business partners and competitors - many of whom have access to greater computing power than we have.

- The role of high performance computing, indeed its importance in a healthy growing economy, was stated consistently and often in the material studied by the Task Force and in the presentations which were made to the group (see Attachment Two). From this information it was evident that without our own, or ready access to, high performance computing, Alberta's industrial/commercial sector will likely not be internationally competitive. Although it may take a while for the lack of availability of high performance computing to be felt, the negative consequences will be ultimately realized and it will be difficult or nearly impossible to retrieve a competitive position. In particular, research and development will fall behind, and our industries, government and universities will likely not be able to hire and retain essential, qualified professional staff capable of meeting world competition. This is particularly important in Alberta as we move from a high dependence on a resource base economy to one that is more diversified and technical in nature.

- Supercomputers can suggest solutions to a number of complex problems in a holistic way. Thus, they can often deal with real world problems without having to make simplifying assumptions, or breaking problems down into their component parts and dealing with each separately, as would be the situation with less powerful computers.

- Several of these types of problems are currently being encountered by Alberta businesses, or will soon be encountered such as in the signal processing of seismic data, petroleum reservoir modelling, drug design, materials design, design of structures, weather and environment modelling, modelling of health trends, educational applications, telecommunications infrastructures, agricultural applications, etc.

- Work on these problems requires supercomputing capability linked into Alberta businesses for commercial exploitation. We have qualified persons in industry and universities to work in this environment.

- There should be a self-sustaining opportunity for Alberta organizations, after the pump has been primed. But, ongoing support, possibly by Government, may be needed to maintain a leading edge facility, without which capability the effectiveness of a high performance computing environment is diminished.

SUPPORTING CONCLUSIONS

Focus

High performance computing, including high speed communications, are required in order to achieve global competitiveness in industrial operations, new products and services, government services, academic research and education in areas such as:
- processing business data;
- signal processing of seismic data;
- hydrocarbon reservoir modelling;
- drug design;
- design of new materials;
- forecasting environmental effects including weather, etc.

The cost of a high performance computing facility is prohibitive for many individual institutions, but the economy of sharing an accessible supercomputing facility would create an advantage for Alberta firms.

The management of the system must respond favorably to the respective research and operational needs of the private sector, government and universities.

To maximize the use of the facility by business, the facility must be managed by a private sector organization. Such an arrangement will probably provide a greater likelihood of selling a newly developed Alberta service outside the Province, than would alternative management arrangements.

Network

An integral part of any high performance computer system is a high speed communications network. Alberta needs such a high speed network. Note: The U.S. system currently has an access rate of 1.5 million bits per second and is being upgraded to 45 million bits per second. It should also be noted that this system is used for much more than data transmission.
Note: The current Alberta system has an access rate of 56,000 bits per second.

Some Alberta companies privately lease 1.5 million bits per second facilities from Telus for their high performance computing transmission requirements.

High speed communications are largely regulated by the Federal Government. The regulators are primarily concerned with existing market conditions and inhibit Canada's and Alberta's global competitiveness because they haven't adequately recognized, as have our competitors, that data communication facilities are the new trade routes for industry.

Human Resources

As pointed out earlier, Alberta will become more reliant in future on the ingenuity of our people to translate ideas into new globally competitive goods and services through the application of technology.

Many of these people, who will be working in numerous diverse fields, will require access to high performance computing power. These fields will be as different as large scale business transactions and cosmology, molecular structure and seismic data processing, drug design and environmental forecasting for example. One of the problems to be faced will be to train sufficient people in information technology, and more specifically in the area of high performance computing, to the level required to meet Alberta's future economic needs.

Research and development in many diverse fields by eminent scientists will be encouraged within our universities through the provision of a high performance computing infrastructure. Both the work of these scientists and the available facilities will attract other researchers and encourage our own students to excel.

Marketing and Awareness

To be successful, a high performance computing centre will have to achieve a high level of participation by business and others, including academia. Given the number of high performance computing professionals in the hydrocarbon and related industries, and academia in Alberta, the potential to achieve this situation is higher in Alberta than in many other jurisdictions in Canada.

Success will require early access to a complete range of high performance computing hardware, including supercomputers.

ATTACHMENT ONE

HIGH PERFORMANCE COMPUTING TASK FORCE MEMBERS

Chairman: Dr. Marshall Williams, Board of Directors, TransAlta Utilities Corporation, Calgary

- Dr. Anthony Harckham, Director of Research and Development, Alberta Government Telephones, Edmonton
- Dr. Bob James, Vice-President (Research), University of Alberta, Edmonton
- Dr. Roy Lindseth, Executive Vice-President, Teknica Petroleum Services Ltd., Calgary
- Mr. Stan Petrica, Executive Director, Computer Processing, Alberta Public Works, Supply and Services, Edmonton
- Mr. Hugh Stanfield, President and CEO, Pulsonic Corporation, Calgary
- Dr. Mike Ward, Vice-President (Research), University of Calgary
- Mr. Walter Yeo, Manager, Shared Computing Services, Shell Canada Resources, Calgary

Secretary: Dr. Alan Vanterpool, Premier's Council on Science and Technology

ATTACHMENT TWO

LIST OF PERSONS INTERVIEWED

Alberta Forestry, Lands and Wildlife	Mr. Mike Toomey, Assistant Deputy Minister, Land Information Services Division (by telephone)
Alberta Research Council	Dr. Ernie Chang, Head, Advanced Computing and Engineering Department
Art Monk and Associates	Mr. A.E.R. Monk, President
Atmospheric Environment Centre, Dorval	Mr. Ian Findleton, Director of the Supercomputer Centre (by telephone)
British Columbia	Mr. Ron Woodward, B.C. Ministry of Advanced Education and Job Training; Mr. Mike Volker, Advanced Systems Institute; Dr. Calvin Shantz, University of B.C. (all by telephone)
Canadian Information Processing Society, Calgary	Ms. Jean Douglas, President (by telephone)
Canadian Institute of Advanced Research	Dr. Fraser Mustard, President, Ms. Dorothy McKinnon, Executive Assistant
Control Data Corporation	Mr. Bob Bolton, Consultant (by telephone)
Cray Research (Canada) Inc.	Mr. Paul Gerritsen, Account Manager, Calgary
FPS Computing	Ms. Pat Comer, Sales Representative (by telephone)
Fujitsu Canada Inc.	Mr. Peter de Merindol, Vice President, Information Systems Group, Toronto; Mr. Ron Crossan, Manager of Large Systems, Toronto
Hitachi Data Systems	Mr. Neil Thompson, Vice President, Western Canada (by telephone)
IBM Inc.	Mr. Troy Wilson, Vice President, High Performance Computing Strategies, Kingston, N.Y.; Mr. Clark Quinton, Vice President and General Manager for Western Canada, Vancouver; Mr. Mike Kirtz, Branch Manager, Calgary; Mr. Brian Olafson, Branch Manager, Edmonton

Industry, Science and Technology, Canada, Ottawa	Mr. Pat Simpson, Director, Technology Applications, Information Technologies Branch (by telephone)
InteCura Consulting Ltd.	Ms. Pat Glenn, President, Edmonton
Jade Simulations International Corporation, Calgary	Dr. Brian Unger, President
Manitoba	Dr. Jim Reichert, Mr. Mel Cornell, Manitoba Industry & Commerce (by telephone)
Natural Sciences and Engineering Research Council	Dr. Pardeep Ahluwalia, Physical and Mathematical Sciences Directorate (by telephone)
Nova Scotia	Ms. Lucia Fanning, Nova Scotia Council of Applied Science and Technology; Dr. Peter Jones, Dalhousie University; Ms. Lynn Tammet, Industry, Science and Technology Canada, Halifax (all by telephone)
Ontario Centre for Large Scale Computing, Toronto	Mr. Bruce Attfield, Executive Director
Saskatchewan	Mr. Wayne McElree, Saskatchewan Economic Diversification and Trade (by telephone)
STM Systems Corp.	Mr. David Parker, Director, Western Region
Technology, Research and Telecommunications	Mr. Peter Noden, Director, Research and Planning
University of Alberta	Dr. Ward Finlay, Department of Mechanical Engineering (by telephone)
University of Alberta	Dr. Monica Beltrametti, Director of Computing Services
University of Calgary	Mr. Alan MacDonald, Director of Libraries
Xerox, Palo Alto Research Centre	Dr. Mark Weiser, Head Computer Science Laboratory (by telephone)

BACKGROUND TO THE REPORT OF THE HIGH PERFORMANCE COMPUTER TASK FORCE

PART TWO

Part Two contains explanatory material which provides support for the material provided in Part One.

Part Two starts with COMPUTERS EXPLAINED, which describes in lay terms how computers operate and what software does (pages 2-10).

Next follows a description of NETWORKS, especially as applied to the Canadian situation (pages 11-13).

This is followed by a section on SUPERCOMPUTERS, what they are used for, by whom, what the benefits are and so on (pages 14-20). This is supplemented by a list of the GRAND CHALLENGES FOR HIGH PERFORMANCE COMPUTING (pages 21-22) and a brief description of HIGH PERFORMANCE COMPUTING IN ALBERTA (pages 23-27), and the University of Calgary's CYBER 205 experience and some of the benefits derived therefrom (pages 28-29).

There is then a section on the UNITED STATES EXPERIENCE (pages 30-32), followed by a list of CANADIAN SUPERCOMPUTING INSTALLATIONS (page 33), and a general bibliography for the report.

COMPUTERS EXPLAINED

The current world of computers is depicted in Figure One, Part One of this report. In the lower left hand corner are the microcomputers, known as personal computers (PC's), laptop computers, or notebook computers in common language. By far the largest number of computer users use these machines - for word processing, accounting, billing, simple mathematical calculations, playing games, keeping record of appointments, etc., for personal, office and small business use.

The mid-range on the figure includes the workstations used by engineers and scientists. Mid-range computers are used not only for scientific and engineering calculations and design work, but also for hotel and airline reservations, accounting, billing and credit card accounts for medium sized firms.

Toward the top end of the figure on the right hand side are the high performance computers. These may do very complex but highly specialized jobs as does a "graphics machine", or very large jobs as does the Alberta Government's "high-end mainframe", or extremely involved calculations using massive amounts of data, as does Shell's Cray computer in Calgary, used for seismic processing.

The most stable characteristic of computers is change. About every three years the number of operations per second for a given cost increases ten times. By the year 2000, today's supercomputer will be no more powerful than a new mini-supercomputer will be then. The Cyber 205 bought in 1986 at the University of Calgary, is comparable to today's high-end mini-supercomputer. By the year 2000, supercomputers will likely be able to handle one trillion operations per second.

A further description of computers and software is given below.

PARTS OF A COMPUTER

- <u>a keyboard</u> - very similar to a typewriter keyboard - is used to give instructions to the computer, often in the form of words, but also in code using symbols such as *, @, /., : and so on - in computerese, the keyboard is used to "talk" to computer.

- <u>a video display</u> - a video screen, but usually one giving a much clearer image than a home TV screen - lets the operator see what is being communicated to the computer, and what the computer is communicating to the operator - the screen may contain words, code numbers, diagrams, animated displays, etc., - in computerese the video display enables the computer to "talk" to the operator.

- <u>a computer processing unit</u> - i.e., the computer in laymen's terms - this is the brains of the system and contains the chip(s) mentioned later.

- volatile memory storage - which may be other silicon chips which are used to temporarily store information and then give it up to be used for further word processing, or calculations, or for video display or to send data to another computer, etc. - volatile memory storage is analogous to the part of your brain you use when you are carrying on a conversation with someone.

- non-volatile storage - which may also be other silicon chips, or more likely magnetic tapes and magnetic or optical disks which can be physically removed from the system if desired - non-volatile storage is analogous to the part of your brain you use when you remember some event of long ago.

- a modem - is a device inserted in the circuitry between a computer and a telephone outlet which converts the computer signals to something which can be transmitted through the telephone system to another modem attached to a computer at a location remote from the computer sending the message.

WHAT IS COMPUTING?

Computing of any type can best be explained by comparing the entire process to electricity and how it is generated and used. Firstly, the computing cycles that are available in anything from a calculator to a supercomputer are analogous to the electrical power supplied from sources ranging from batteries to a major power generating plant. Both the computing cycles and the electrical energy are commodities that are relatively useless by themselves but are essential to power a vast array of products.

There are many parallels between computing cycles and electrical cycles. If the power from a "C" cell battery was comparable to the computing cycles from a calculator then a personal computer would be like a car battery. Workstation computers would compare to portable "Honda" generators while parallel processing computers might be thought of in the same way as the modern windmill arrays used for power generation in several locations. In this later case, both technologies still have some problems to be overcome. At the upper end of the scale, coal and oil fired generating plants have similar characteristics to the large mainframe type of computers while the massive hydroelectric facilities would be comparable to the supercomputer.

When properly applied, the power from a small battery is no less appropriate than the current available from a power plant. The battery will do a superb job of powering a portable radio but is totally inadequate for running a large broadcasting station. It may not be practical to use a supercomputer to solve a simple problem that can be handled by a calculator but it would be futile to attempt massive computing projects on a tiny device.

Another similarity between computing and electricity is the distribution requirement at the high end. A power generating plant is only useful if it has the transmission lines necessary to connect many consumers with the electricity generated. This is not a consideration for battery power and is of little concern for small, dedicated generators. In much the same way, a major computing installation must be able to deliver its resources to the customer over fast, efficient communication networks.

It is essential in both cases to match the cycles produced to the requirements of the user community. Standards are critically important to both technologies. If a new power utility went into production today, generating electricity at 50 Hz, its output would be almost useless in the 60 Hz North American market. People who adapted their equipment to this proprietary type of power would find themselves locked into one source regardless of what the supplier decided to charge. If they built machines using this form of power, their market would be restricted to those other individuals confined in the same situation. The parallel between the kind of electricity produced and the type of computing cycles generated is to be found in the system operating software. A computer utility that supplies a non-standard operating system is no better than the 50 Hz power company. In the early days of electricity, even as the present standards were emerging, many forms of power existed and different groups promoted their own product. Today, the emerging standard in computer operating systems is UNIX and it is apparent that these are the computing cycles that will be in demand.

HARDWARE

The basic element of all computers is the chip. Commonly produced from silicon, these miniaturized electronic circuits contain thousands of simple switches that can be set to one of two positions; on or off. Depending on its status, the switch is considered to be either 0 (zero) or 1 (one). Fundamentally, this is all a computer knows. Every computer performs every task using only this binary number system. Letters and symbols are represented by specific codes. These zeros and ones are the bits in computer language. Eight bits together are called a byte. Computer words are the standard units handled by a machine and lengths vary depending on the particular hardware. Personal computers of the 286 variety use a word length of 16 bits which means that after reserving one bit for designating the sign (+ or -) the largest number that can be represented is 32,768 (2 to the 15th power). If the desired number was 32,769 the computer would have to employ two words thus adding to system overhead. Mid-range computers often use a 32 bit word length while large mainframes and supercomputers are generally 64 bit machines.

The functions that a chip can perform are determined primarily by the circuit design etched into the silicon. Some chips are used only as computer memory and perform no calculations while others are intended to carry out only certain types of instructions. Large scale computers generally incorporate a variety of limited purpose chips installed on complex circuit boards. Smaller machines such as workstations will utilize a single microprocessor chip in which almost every circuit element has been imbedded in the silicon. How efficiently tasks are performed at the chip level depends to a large extent on the manufacturing process and the ability to put more circuitry and gates on a single chip. Some chips currently being produced have been compared in their performance to the original Cray I supercomputer.

Gaining increased computer performance through better chip design and improved manufacturing techniques is beginning to encounter very real physical limits. The ability to pack more circuitry into less space will require considerable research. New materials such as gallium arsenide are being employed to make chips for several advanced machines. These exotic compounds are used in place of silicon due to their ability to move signals quickly and switch their gates faster. It has been estimated that, for the foreseeable future, these efforts will only yield a doubling of speed every three to five years.

There is one important dimension in computing that is often overlooked and not well appreciated. This is the element of size. A good analogy would be to compare a Lear jet and a 747 Jumbo. The Lear can do everything the 747 can do and even several it cannot, such as land on short runways. They will both move their passenger load to a destination at about the same speed but the 747 is capable of delivering fifty times more people along with baggage and even freight. Of course, if only eight people are travelling it would make more sense to use the smaller aircraft if there were only the two choices. Expanding this analogy would correlate a high-end personal computer with a single seat jet trainer and supercomputers would range from the Concorde to the Space Shuttle.

COMPUTER CLASSIFICATION

All computers in use today can be grouped into one of three categories based on their method of solving problems. These classifications are scalar, vector and parallel. Within each of these groups there are subclassifications with varying degrees of complexity but the fundamental problem solving techniques of the sub-groups do not change appreciably. Using this method of evaluating computers tends to classify workstation machines together with large mainframe devices but this can be shown to be accurate and reasonable from a computational standpoint. The significant differences between a workstation and a mainframe are not in how the computing is performed but in other factors such as memory capacity, data input and output facilities (data I/O) and system software. It is these differences that allow the mainframe to function like the Jumbo jet while the workstation is analogous to the Lear.

SCALAR COMPUTERS

The scalar type of computer is the general purpose machine. It can be programmed to perform tasks ranging from word processing and data base management to complex scientific instructions. Often referred to as the central processing unit or CPU, a scalar processor must be present in every type of computer architecture. This category of computer will frequently have special accelerator boards or chip sets installed to expedite specific tasks such as math functions or graphics rendering but they are not essential to the scalar operation.

The functioning of a scalar computer is a very linear process that can be compared to a person using a hand calculator equipped with memories. Firstly, it is determined where the input data is located and a button is pressed to retrieve this information. Then it is decided which function will be performed on the data and this selection is made. Next, the appropriate operand is recalled from memory and the "equal" button is pressed to obtain the result. Finally, the answer is output to the display and may be stored in some memory location. In this example, the program controlling the operation of the machine is running in the person's mind. There are many programmable calculators on the market today where complex sequences can be loaded and run repetitively with varying input data.

However, scalar computers are capable of much more than the functions normally found in a calculator. Their circuitry and built-in instruction sets make it possible for programmers to develop complex routines that can be used to perform logical operations such as controlling other machines. For many applications, a straightforward scalar computer is the best choice and for some jobs, it is the only choice at the present time.

VECTOR COMPUTERS

Vector computing is a term that is somewhat misleading. The devices in common use today are more correctly referred to as integrated scalar, vector computers. As noted above, the scalar element is an essential part of this style of computer. All of the administrative work is carried out by the scalar device which also controls the operation of the special vector unit or units.

A vector is nothing more than a string of numbers that all have to go through one or more operations. The vector processor is a specialized piece of hardware that can be compared to a production line in a factory. The scalar component is the equivalent of the foreman or supervisor. A program running in the scalar processor instructs the vector unit as to what sequence of operations will be required for the project. When the vector unit signals that it is ready to commence working, the scalar unit initiates the data flow from the appropriate locations in memory and posts instructions about where to place the finished products. At this point, the scalar becomes relatively idle and is able to start another job if one is available.

The production line set up in the vector processor is referred to as a pipe. Each element of the vector is processed through the various stages defined in the initial setup procedure. Once the pipe is filled, processing proceeds at a very rapid pace when compared to the "one-step-at-a-time" approach that the scalar processor would use. Generally, the longer the vectors and the more complex the process the greater will be the advantage of this type of computing. Several machines of this architecture incorporate more than one pipe in the vector unit to allow for more than one process stream on a single data set or to allow several jobs to run concurrently.

PARALLEL COMPUTERS

Parallel processing has become a vast subject that on one hand is simple in concept yet on the other hand is complex in implementation. The basic idea is that some quantity of separate processing units can be connected together electronically and through software instructions they are made to share the computing workload on a single task. Systems range from just a few elements (two to eight processors) up to complex arrays of many thousands of units. Major divisions in architecture are developing and each different approach seems to exhibit great potential on a certain spectrum of problem types. At present, no one design seems capable of operating efficiently across a range of performance to the extent that it could be deemed a multipurpose computer.

Some of the problems being encountered are hardware related but these are gradually being resolved. The major complexity has been to find a universal method for effectively splitting up a task and distributing the problem along with the essential data to the separate elements in the computer and then bringing everything back together at the end. This is primarily a matter of software development. If a particular program has been specifically designed and written for this type of architecture, the results can be quite dramatic for certain kinds of problems and data sets.

Parallel computing holds the greatest promise for dramatic increases in processing speed for the future. The barriers to achieving very high performance over a broad range of applications are not so much a question of hardware as they are of software. It is neither feasible nor desirable to rewrite all the millions of lines of code that presently exist in order to fit a parallel architecture. The answer is to be found in developing new software technology that will exploit the inherent potential of this type of computer yet allow conventional, existing programs to run.

An Alberta company, Myrias Research Corporation, developed a "moderately" parallel high performance computer with considerable financial assistance from the Alberta, Canadian and U.S. Governments. Their machines are installed in over half a dozen locations in Canada and the U.S.A. The software is more successful than that of many other emerging parallel computing systems, and the machines are apparently relatively easy to use on several of the applications on which they have been tried. Although the original company went into receivership, Myrias' technology was purchased by DATEK Industries Ltd. of Edmonton and further development is being continued by their subsidiary, Myrias Computer Technologies Inc.

In spite of the formidable difficulties in developing a general purpose massively parallel supercomputer, such machines are clearly part of the high performance computing scene today, and will be much more common by the end of the decade.

SOFTWARE

Two major categories of software exist. The first type is most frequently referred to in a collective sense as operating system software. This complex combination of programs provides the life blood and basic motor functions for the hardware and also provides the means with which a user can begin communicating with the device. Included in this category of software are numerous programs of differing types that allow programmers to do their job effectively. The operating system software provides the "personality" of the combined elements and when it has been installed in the hardware the combination may properly be referred to as a computer system. Any other use of this term is mis-information which has become common practice in many circles.

The second level of software consists of the application programs. These are the detailed instructions that are provided to the computer system in order to accomplish the specific tasks that justify their existence. Operating software supplies the ability to do the job while application software provides the job itself along with the directions on how it should be accomplished. From the simplest problem of adding two numbers together to complex excursions into nuclear physics all application programs share three major elements:

- what data to use and where to get it (load)
- what function(s) to perform on the data (compute)
- where to place the result upon completion (store)

Without the coordinated efforts of these two types of software, every computer is incapable of performing even the simplest task. Certainly the design and quality of the hardware plays a key role in the eventual performance of the complete computer system but it is generally recognized that software development has achieved a level of importance equal to or even surpassing that of hardware design. In fact, as noted in the section on parallel computers, these architectures that show great promise cannot yet realize their full potential due to a lack of adequate software.

The overall field of developing system software ranks in complexity with the design of the hardware systems themselves. Some of the programmed instructions are imbedded directly in the chips or otherwise incorporated into the electronic circuitry. This tedious work is carried out by a relatively few programmers in the world using complex programming tools with names such as machine language, assembler code and micro-code. In terms of human physiology, this level of system, software could be compared to the autonomic functions that control things like our heartbeat.

As the system functions move further away from the nucleus of the computer, the complexity of the programming and the tools used may decrease but it is still very demanding work. Building the systems software necessary for a computer to be able to input and output data may be compared to teaching an infant how to acquire information through reading and various other means and to communicate that knowledge fluently and flawlessly. Development of powerful system software remains one of the most challenging frontiers in the field of computing.

Almost all programmers who write application software use a high level computer language to develop their products. Scientists and engineers generally use languages such as FORTRAN or "C" while the most popular for accounting and business use is COBOL. These languages have specific grammatical structures and rules, just like any spoken language, that allow a person to put their ideas and formulae into a logical structure or "story". However, a computer will not be able to understand this program directly because, as previously noted, it can only comprehend things that have been reduced to a series of switch settings representing zero or one.

Specific to every brand of computer and each different language there is a specialized piece of software which runs as part of the operating system called a compiler. The compiler translates the entire program written in the high level language which is called source code into instructions that can be understood and implemented by the computer or something called binary code. "Binary" simply means "two" and refers to the zero and one number system that a computer understands. As an example, the decimal number sequence - 0, 1, 2, 3, 4, 5, 6 - would be translated by the compiler to the binary code - 0, 1, 10, 11, 100, 101, 110 -. The binary representation for the letter "e" would be 1100101.

Compiler technology is one of the most significant areas of research in the computer industry today. Apart from the purely mechanical translation of information, advanced compilers are able to examine the source code program for its logical intent and suggest ways to reorder data and processes to increase efficiency. Both parallel computers and vector computers rely on being able to organize the data and the problem to be solved in a certain way in order to achieve their enhanced speed. This data organization could be done explicitly by the programmer in the course of developing the program or it can be done implicitly by employing these special types of compilers. The most efficient way to write a program from a programmer's perspective is generally to assume the environment will be a scalar computer. Creating a program to run on another particular type of architecture can be very difficult and time consuming. Often, the program created by these special efforts is very restricted as to the type of computer on which it will run.

Vector computers rely on the data and problem being organized in the linear fashion described in the hardware section. The types of programs that will do this job are referred to as automatic vectorizing compilers. Parallel computers require that the data and problem be broken down into discrete pieces that can each be handled by a separate

processor. Software that can perform this function is called an automatic parallelizing compiler. Primarily because vector machines have been around longer than parallel systems, vectorizing compilers are more advanced than the parallelizing systems. Several of the more advanced automatic vectorizing products can often reach levels of efficiency greater than the average programmer can achieve using laborious manual techniques. Automatic parallelizing systems are still largely at the prototype stage of development and success rates fluctuate widely.

SUMMARY-WORLD OF COMPUTERS

Computers have quickly entrenched themselves in our modern society to the extent that they have become indispensable yet they are still very much misunderstood. The rapid evolution of the technology combined with the amazing feats the present machines can accomplish seem to make these devices even more incomprehensible. The foregoing overview has left out large amounts of detail but it does discuss most of the information required for a good understanding of the subject as it is at the present time.

In the not too distant future, other dramatic developments will take place that will make our current systems appear as clumsy and obsolete as the old Eniac machine built from vacuum tubes back in the 1950's. Even now research is underway and prototype systems have been built that attempt to emulate the neural networks of the human brain. Commercial software that can simulate this neural process in existing computers has already been developed by Alberta companies.

It is difficult to compare any other human endeavor to the invention and development of the computer. Aviation history has been relatively short and very dramatic but much of what we have today would not be possible if not for the computers that were used to design, test and operate the machines over the past thirty years. The ability to reason has been said to be the most important factor for the human race since the dawn of mankind. It has also been said that the most important contribution of computers to society is that they enhance our ability to reason. If these two statements are deemed to be correct, then logic would dictate that we make maximum use of the most advanced computer technology available and strive to create even better systems.

NETWORKS

Just as individuals communicate through the telephone network, so do computers. Computers however make more demands on the telephone system than do individuals, especially with respect to speed and reliability. For example, an interruption of a tenth of a second when two individuals are talking would not be noticed. However, such an interruption when data is being transmitted between two computers would disrupt the entire communication.

NETWORK MANAGEMENT

While telephone companies actually operate these networks, high-speed networks are often managed by other organizations. As an example, the National Research and Education Network (NREN) in the U.S.A. is managed by Merit Network, Inc. In this case, Merit acts as a standard setter, fee collector, and the payer of the basic network costs.

In British Columbia and Quebec, the provincial governments subsidize their network managers so that the cost of the services to the users is less than the tariff cost to the manager.

NETWORK ACCESS RATES

One important consideration for users is the "access rate" which is measured in bits per second (bps):

- in Canada and Alberta, CA*net operates at 56,000 bps;
- in the U.S. the comparable system, NREN, operates at T1 (or 1.5 million bps) and is being upgraded to T3 (or 45 million bps);
- some short distance systems operate at FDDI rates (or 100 million bps), such as the proposed network backbone within the Universities of Alberta and Calgary.

ISTCNET

Industry, Science and Technology Canada is proposing a national network, to be known as ISTCNet, at a T1 access rate, for 1994. It is estimated to cost $60 million of which the Federal Government would pay $30 million, the private sector $12 million, and the provincial governments collectively, $18 million. Alberta's share would be under $800,000. In addition, Alberta would be expected to provide a T1 link between Edmonton and Calgary.

NETWORK PRICING

AGT Limited is now subject to CRTC, as will be all the telephone companies within Telecom Canada. One of the regulatory provisions of CRTC is that no discounting shall be allowed: all customers pay the same price for the same service whether they use one unit of service or many units.

To provide a subsidized service some network management organization has to act as a reseller (as noted for B.C. and Quebec above). This organization pays for phone lines at the standard rates but can charge back to users in any way that they like. Such organizations are subsidized either through provincial or federal governments.

Pricing in Alberta for T1 between Edmonton and Calgary is:

1. $9,000 one time construction charge

2. $550/month access charge

3. $12,000/month carriage charge.

Using Bell's filed rates for T3 the expected charges between Edmonton and Calgary might be:

1. $9,000 one time construction charge

2. $4,400/month access charge

3. $108,000/month carriage charge.

In Alberta access to ARnet, managed by the Alberta Research Council is $28,000 annually. ARnet provides access to CA*net and NREN.

COMMENTS ON THE NETWORK PRICING SITUATION

Communication systems for computers are now catching up with the advances made in the processing technology. However, the tariff structure for data communications in Canada is woefully inadequate. There seems to be a lack of understanding of the significance to the national economy that high speed, cost effective data communications could have. Canadians currently pay several times the rate charged for equivalent services in the U.S.A. The rationale seems to be that if one request for data transmission requires a long distance circuit that could carry 1000 personal phone conversations then the charge will be equal to 1000 phone calls. The logic appears simple but there are numerous other factors that must be taken into consideration. Although tariff rates are a Federal jurisdiction, the existing situation will have a dramatic effect on any Provincial initiative related to information technology, including a supercomputer installation.

The basic issue is that it is very costly to ship data rapidly. Ways must be found to avoid this problem in Canada, as has been done in the jurisdictions with which we compete.

It would not be worthwhile installing a supercomputer in either Calgary or Edmonton unless there were a T1, and preferably a T3 link between the two cities. This must be a major consideration in any government decision. One significant characteristic of a supercomputer is its ability to quickly process large volumes of data. Using existing communication facilities to provide data to the machine would be akin to trying to fill an oil tanker with a garden hose.

Substantial opportunities for international business will be available to the society that quickly realizes that its communication facilities are its trade routes and that its computing installations are its manufacturing plants. It may be pleasant to reach out and touch someone but it is much more profitable to reach out and process their data.

SUPERCOMPUTERS

HIGH PERFORMANCE COMPUTING AND SUPERCOMPUTING

The terms "high performance computing" (hpc) and "supercomputing" (S/C) are often used interchangeably but this usage is not correct. It is said that the hpc term was originally introduced by a large computer manufacturer that had powerful machines but lacked equipment in the supercomputer category.

The generally accepted view is that hpc describes an overall environment whereas S/C is a specific type of operation. High performance computing is very subjective and may be applied to many different situations but supercomputing can only occur when a machine of the supercomputer class is being utilized. An environment that contains a supercomputer is likely to warrant the hpc designation but the term has also been applied to operations that contain only workstations or even mainframe type accounting computers. Recently, personal computers that utilize the Intel 486 chip have been designated as hpc machines by people who are focussed on the PC environment.

To most computing professionals, high performance computing generally implies a coordinated utilization of a variety of computer types in a network environment. Ideally, the systems would range from workstations and specialty graphics rendering computers all the way up to a supercomputer. Each element in the network would be assigned to the tasks that it performs most efficiently.

Many hardware manufacturers today are choosing to call their machines supercomputers but the title rightfully belongs to only a few products at any point in time. There are presently no rules or laws controlling the use of the supercomputer term. Like other titles that are awarded based on relative performance, the right to be designated a true supercomputer must be continuously defended or it will be lost. A former boxing champion is analogous to a former supercomputer. Each may still be very good at what it does but others have taken over the title. The Cyber 205 falls in the "former" category of supercomputing.

Speed alone does not define a supercomputer. In addition to being able to perform calculations at speeds near the peak of current industry expectations, the machine must also have other minimum characteristics. In order to qualify for serious consideration in this category by the majority of potential buyers a computer must:

- have a powerful (preferably non-proprietary) operating system
- provide high quality, reliable programming languages and compilers
- have a very large memory capacity with fast access
- be able to move data (I/O) in proportion to its computer speed
- be capable of multi-tasking and high project throughput
- be capable of supporting a large number of peripheral devices
- be "general purpose" to accepted industry standards
- utilize existing and emerging standards, protocols, languages, etc.

- provide a hospitable environment for the user community
- support a wide range of libraries and standard application programs
- offer an evolutionary growth path to avoid obsolescence

A formulae I race car can lay claim to many outstanding feats of performance but it has limited application in the off-track world. If the above list were translated into comparable automotive terms, the race car would fail in every category; but it is certainly fast. The same or similar situations exist in the realm of supercomputing. It too has its formulae I candidates. The obvious factor with these machines is that they are designed and built for a specific purpose. They can do an excellent job within a narrow range of functions but they cannot be considered "general purpose". Much of the knowledge gained in building such devices may eventually benefit consumers but there are limited uses for the machines themselves. Great care must be exercised in the selection of a supercomputer that is intended to service a large, diverse user community.

MORE ON SUPERCOMPUTING

Supercomputing is one of the leading edges of computing technology. It is employed to address problems which are large scale in the quantity of computer processor time required, or in the quantity of data to be handled. It is founded on the availability of high performance computer hardware. The growth path for it is seen as one of increasing parallelism and complexity of internal organization, though raw processor capacity is far from exhausted.

The different threads that can be distinguished, scalar, vector and connectionist, all have specific application domains where their individual characteristics provide greatest benefits. Supercomputing is thus a quality of operating at a leading edge, not of being a certain implementation, nor of simply being a tool to address numerically intensive problems.

Just as personal computers required software to make them widely used tools, so do supercomputers require the interface to carry the capabilities of the machines into delivery of useful results on specific applications.

The hardware developments and operating systems are driven out of the U.S. and Japan. The presentations by Cray, IBM and Fujitsu to the Task Force, indicated a commitment to continuing the development of the hardware: all these organizations believe that there is a growth path in this area with existing and foreseeable technology, which will carry supercomputers for at least 10 years, with comparable rates of performance increase to those experienced in the past.

The software tools for using supercomputers are becoming more standardized: UNIX has emerged as the preferred operating system. It may be expected that with this degree of standardization more attention will be paid to producing improved programming languages and systems. Such capabilities would shorten the production time for software, would increase the reliability of the produced code, would allow more complex problems to be handled and would enhance the transferability of code into production through improved maintainability.

Parallel developments may also be expected in the area of massive data storage, with standards being set for data structures and access methods which could result in the creation of data stores independent of the actual processing capabilities but tightly coupled to them.

There are many operational issues which need to be satisfactorily resolved for commercial users to treat supercomputing as normal: these include security, reliability and accessibility.

Like any other facility, supercomputing depends on being linked to its potential user community, through reliable, high-quality access paths, i.e., a network, through the availability of long-lived, standardized tools, and through a culture of use. This latter element in turn depends on an agreed value for the facility, accepted purposes, informed users and an educational system which provides growth and continuity for these.

COMMERCIAL vs. SCIENTIFIC COMPUTING

There are two distinctly different kinds of computing. One is commonly referred to as "commercial data processing" while the other is frequently called "scientific data processing". Neither of these terms is completely appropriate or adequate to describe the functions involved, so some explanation is required. The implication of using the "commercial" and "scientific" designations is that scientific work is all research with no commercial application, which is completely false.

Commercial data processing is the type of activity carried out by banks, insurance companies and corporate controllers for example. It is the process that occurs when an automatic banking machine is accessed or a purchase is made at a department store. Credit cards, electronic mail and library systems also fall into this category. The computing work does not require large amounts of actual calculation of numbers. It is primarily a more efficient means of doing the work previously handled by tellers, accountants and file clerks. The computers that do this work are powerful but they are oriented toward storing and retrieving vast quantities of data very quickly with the occasional short diversion to produce a monthly utility bill or financial statement. Although it is low profile and relatively unsophisticated, this type of computing represents the major use of computers in our society today.

Scientific data processing is also frequently called "numerically intensive computing" or NIC. The second name is more accurate and descriptive of what is actually taking place. This is the type of computing that is done by oil companies for reservoir analysis, seismic companies for hydrocarbon exploration and university researchers in many disciplines. It is also used extensively in such areas as medical research, pharmaceutical drug design, meteorology, advanced materials development and environmental modelling, to name a few. NIC, using large machines, could be used to predict the environmental impact of a pulp mill installation and in many labs it is already employed as a powerful tool in the fight against AIDS. The principal characteristic of this type of computing is the extensive, complex calculations that are performed on large quantities of numerical data. Depending on the magnitude of the project, this work is carried out on machines ranging from workstations to supercomputers.

With the appropriate software installed, a personal computer can do either computing function quite well as long as its moderate capability is not exceeded. However, as the respective demands escalate, computer systems and software designed to perform one type of computing become less proficient at being able to handle any substantial amount of the other type of work. Attempts are under way to produce systems that can handle both functions efficiently and cost effectively but any success would appear to be several years in the future. This timing will correspond to the convergence of many disparate aspects of the entire computing industry.

THE USE OF SUPERCOMPUTERS

Supercomputers are used for addressing very large scale problems. Essentially the common set of "grand challenges" all are based on the application of a detailed understanding of small scale processes to macro problems. For example, the cosmological challenge is to understand the development of the universe, the largest structure known to man, in terms of the interactions of elementary particles, the smallest divisions of matter and energy that have been so far conceived. The results appear in two ways: discrepancies between theory and observation are resolved either by modifying the view of the way in which the individual interactions are collectively organized, i.e., a change to the growth model of the universe, or by modifying the form of the elementary particle interactions. The supercomputer can thus develop both microscopic and macroscopic theory. Commercially, supercomputers are used in many areas of information technology, where vast quantities of data have to be converted into usable information, particularly where timing is critical.

The supercomputer is a general purpose facility: its cost is kept low, comparatively, though the range of uses to which it may be put. Its particular uses are generated through software which is a lower cost way of accomplishing specific function, as perceived at all levels of use, from the single chip which may control either a CD player, a TV or a microwave oven, through the general purpose personal computer which can maintain accounts, perform word processing or act as a graphic workstation.

BENEFITS OF USING SUPERCOMPUTERS

Two types of benefits have been identified:
- (a) scale of problem that can be addressed
- (b) efficiency of research process.

There are many business processes for which judgement or rule of thumb is the guiding principle. As supercomputers grow in capability some of these informal methods may be replaced by systematic supercomputer models. Examples of this include stock market performance where adequate performance models would allow speculators to lower their risks in the short-term; in the longer term as the models become more prevalent the efficiency of the market could be improved. A side benefit of this would be the early detection of fraud. The first market to become efficient in this way will gain a lead in global trading.

Alberta interest which could be supported through supercomputers include:
1. biotechnology industry
2. materials industry
3. medical research
4. resource industry.

Government itself could become more efficient and more effective through the use of models on which to assess existing and proposed services and support. Areas such as the modelling of health trends, education, weather modelling, environmental forecasting, agricultural modelling, and telecommunications infrastructure come to mind, for example.

The overall effect of supercomputers is to move from a plant-intensive environment to an information and software intensive environment. If a concerted thrust is made in this direction, then this will result in the substantial growth in the software and information industries and probably a relative decline in the manufacturing and assembly industries associated with plant. It will also call for a new specialization, namely of realizing physical implementations of successful, optimized models developed on computers.

This is a realignment which is occurring on a world scale anyway. The question is whether it is the desire of the Province to become an early-adopter.

The natural advantage that Alberta has is familiarity with large scale computing through the resource industry which to date has been a leading user. Another advantage is the relatively limited investment in manufacturing plant which would be displaced by this move.

The costs incurred would be the development of the infrastructure, including the educational one, to make such a move possible and economic for business earlier than it would be without such governmental involvement.

The industry which develops has strong export capabioities, and may, through communication networks, ship its product anywhere in the world within very short time spans.

HOW IS SUPERCOMPUTING ACCESSED?

Supercomputers are expensive facilities in terms of investment and true operating cost. Because of the limited lifetime of the hardware, as remaining at the leading edge, there is a heavy investment cycle to maintain.

Although some businesses may possess their own supercomputers, particularly during the build-up phase of the culture of use, there is likely to be a predominance of shared facilities. Even within a single business, there is likely to be a requirement to use the facility from a number of geographically separated locations.

A key part of a supercomputing facility is a high speed network. Supercomputers are largely used through graphic workstations, which assemble data into images which can be readily interpreted. The network has to allow large volumes of data to be shipped from the central facility to a workstation. This network is commonly expected to be at least T1 (1.5 million bits/second) initially and later T3 (45 million bits/second) or beyond.

As mentioned supercomputers are of at least three main types currently: in addition what a supercomputer can be used for is influenced by software available, which is, at present at least, dependent on the community of users accessing a particular supercomputer. The network for access is therefore required to connect to supercomputers on a national and an international basis.

Data capture and data access are also key parts of supercomputer usage. Data repositories capable of managing multi-terabytes of data are required and these volumes of data will need to be shipped to the processing centres.

If supercomputers are to be used for their potential they must deliver that potential to all points of use.

IMPLICATIONS OF SUPERCOMPUTING

Information Age

Supercomputers are becoming a part of the economy: the machines will displace certain types of cost, will create new kinds of value and will be essential for addressing the customization and personalization of product and service which will characterize the economy in the next few years.

The investment in supercomputing is thus a commitment to staying at the leading edge.

Usability

Supercomputers have immense potential: for this to be unlocked it is necessary to develop:

1. understanding of how to use them effectively.
2. development of software tools and methods.
3. standardization.
4. interfaces for users, e.g., for visualization, possibly using artificial intelligence.
5. interfaces between computers and networking to operate at very high speeds efficiently.
6. linkages between research and industry.
7. an affordable network.

Charging

If businesses are using supercomputing facilities to obtain real solutions to real problems then payment will cover the costs of the facility if it is the right tool and is loaded to a viable level. In many industrial situations it can be shown that the most cost/time effective means of solving a problem is to buy supercomputer cycles from a facility.

The provision of facilities for research can be free if they are taken as part of the cost of developing solutions for industrial problems. In such a mode the supercomputer is a business operation. The implication of this is that research would only be undertaken to address known problem areas of actual businesses.

Academic research would have to be paid for as at present through research grants at either a provincial or federal level. It might be expected that with more requirements for computing time, and correspondingly fewer for plant, the costs of doing this research should diminish.

THE GRAND CHALLENGES FOR HIGH PERFORMANCE COMPUTING

A number of areas of study have been identified as "Grand Challenges" for high performance computing. It is recognized that computational methods may not be the only ways in which these "Grand Challenges" will be resolved, but at present the computational route is either being extensively utilized or is the only proposed path.

The overall characteristic of the computational method is to apply some underlying theory repeatedly for very simple elements of the total problem and then to integrate the results through processes which are either considered to be models of the higher level interactions which occur in the real system, or else are stable enough to allow controlled feedback loops to be established.

Two major classes of process emerge: those which are data limited and those which are processing limited. The two categories suggest that different systems may be more fitted to each category rather than a single type of general purpose high-performance computer. Each category may be further broken down, e.g., the data limited group may depend either on large volumes of stored data which is to be analyzed, as in the Mars overflight case, for example. Or on the real time capture of large volumes of data as in the stock market problem, for example.

Some of the "Grand Challenges" are listed in the following table, and suggestions made as to where potential solutions might be used in the commercial world.

Grand Challenge	Manufacturing Industry	Service Industry	Knowledge Industry
Weather Predictions, etc.	Agriculture	Tourism	Regional Planning
Materials Science	Chemicals, metallurgy	-	Design, Architecture
Semiconductor Design	Electronic components, Consumer electronics	Software	-
Superconductivity	Electronic components, Transportation equipment manufacture	Transportation	Design of vehicles
Structural biology	Pharmaceutical	Health Care	Preventive medicine
Design of drugs	Pharmaceutical	Health Care	-
Human genome	-	Health Care	Genetic defect elimination, genetic enhancement

Grand Challenge	Manufacturing Industry	Service Industry	Knowledge Industry
Quantum Chromodynamics	-	Education	Philosophy
Astronomy	Indirectly manufacturers of radio telescopes	Education	Philosophy
Fluid Flow	Vehicles manufacture, Large plant manufacture (power generators); chemical plants, oil refineries	Utilities, Transportation	Consulting
Nuclear Fusion	Indirectly manufacturers of fusion reactors	Electrical generation	-
Combustion systems	Refining, power generation, automobile manufacturer	Environmental monitoring	Design of plant
Enhanced oil and gas recovery	Resource industry, indirectly manufacturers of equipment	Exploration/ production	Consulting
Speech	Communications, Information technology	All services that may be accessed by speech	-
Vision	Manufacturing plant	Security	-

The future will based be on knowledge intensive goods and service producing industries. Alberta's challenge will be to link the "Grand Challenges" to sectoral strengths we now have or can develop. Much of the knowledge base envisaged in the above table is the ability to design devices which have high performance, low cost and low impact on their environment. If Alberta generates the strength to design these systems, and to deploy them first in the Province, we should be able to develop a sustainable economy.

HIGH PERFORMANCE COMPUTING IN ALBERTA

COMPUTER MANUFACTURING

The Task Force has reviewed several documents discussing the importance of high performance computing in both the U.S.A. and the EEC. Reference is made in those proposals to building a strong hardware manufacturing segment to stay competitive with other world powers including the Japanese. Considering the intense competitiveness that already exists in this area, along with the massive investment required, it would appear that building new generations of computers is not where Canada or Alberta should focus its attention. Instead, we should concentrate our efforts to become world leaders in software development and efficient utilizers of the available technology. The article entitled "The Computerless Computer Company" (Harvard Business Review, July-August 1991, pp. 69-80) is an excellent review of this subject. It is significant to note that Microsoft, a company based on software development, has a market valuation that is 70% greater than Digital Equipment Corp., the third largest computer manufacturer in the world.

COMPUTING EVOLUTION

The oil industry began to use computers extensively for various purposes in the 1960's. Geophysicists found that they were the ideal tool for handling their large volumes of data and increasingly complex calculations. For many years, the numerically intensive computer (NIC) industry grew in a long series of revolutionary jumps while communications technology barely moved at all. New computers were introduced complete with their own proprietary operating systems and each product promised a significant advantage over the previous technology leader. With each new machine that had potential to serve in the oil industry another new contracting company would be created to capitalize on the competitive hardware advantage. Firms that were already established had coupled their software technology to the architecture of their then current machine and the lack of standards made it difficult for them to change quickly.

Thus, in the 20 year period from the mid-1960's to the mid-1980's a large NIC computing usage was established in Alberta. The lack of adequate or cost effective communication networks was one of the factors that foiled several attempts to move the many separate players to one large computer platform. Each company stretched its resources to the limit to acquire their own computers that were capable of handling the peak loads of a very cyclic business. Proprietary operating systems remained the norm for each different computer vendor but a certain degree of stability was achieved within product lines. Evolutionary developments were beginning to replace the old revolutionary methods of computer design.

The steady growth of NIC users both in size and technical complexity was only made possible by the influx of a dynamic work force consisting of scientists, programmers, computer experts, technicians and a host of other talented people. Many of these individuals were entrepreneurial by nature and have been instrumental in launching new, diversified ventures which have helped to broaden the base of the Alberta economy.

COMPUTING IN ALBERTA

Any discussion on computing should be prefaced by defining which of the two major types is being dealt with. Even though "commercial" computing occupies most of the hardware resources, it is the "scientific" machines and applications that get most of the publicity. The reason for this is quite simple. Commercial computing deals with the mundane issues of processing the daily events of society and recording the data for reference purposes. Scientific computing deals with the future and helps define the events and technologies that will shape our lives. For newsworthiness, it is like comparing a tax auditor and an astronaut.

Commercial installations will continue to flourish and grow at a rate corresponding to the growth of that particular user community. The types of functions performed by commercial machines are not subject to the same competitive pressures as the scientific community faces. To paraphrase a well known writer, commercial processing is comparable to painting a barn while scientific work is like painting a Picasso. The comments that follow pertain to the field of scientific or NIC computing.

It has been said that Calgary and even Alberta as a whole has one of the highest ratios of computing capacity to population in Canada and possibly North America. This is especially true for NIC computing, but it is also misleading in many ways. Most of our computing capability is generated by numerous small to medium sized facilities in complete isolation from one another. These are like factories generating electrical power, sufficient for their own needs, by using motor generators. The productivity of the factory is governed by the maximum power they can generate. Conversely, when the factory is not busy, the electrical potential goes to waste yet the capital cost of the generating plant remains fixed.

Viewed as a whole, the NIC computing industry in Alberta has a tremendous investment in a large variety of incompatible machines but these are growing obsolete. Most of this equipment was originally installed to service the needs of a dynamic, local oil and gas community that has since begun to deteriorate. At one time, the combined hardware plus the community of people employed to utilize it made Calgary the third largest seismic processing centre in the world. Approximately seventeen percent of the total seismic exploration effort in the world, excluding Russia and China, was processed here. Now, without the ability to access NIC equipment that meets world class standards, this particular industry is faced with a weak domestic market and difficulty in attracting international work.

The Task Force has heard from several sources that Alberta will start to lose prominent researchers from the academic community unless their requirements for access to supercomputing facilities are resolved. The work they are doing cannot be continued without this essential tool. A similar situation exists in the industrial sector. During the expansion period for the oil and gas business, Alberta attracted many talented people to work in our booming technology industry. With reduced activity

by the oil companies in the local market, we are in danger of losing a valuable resource; our highly trained computer work force. There is a burgeoning, worldwide demand for people who are skilled in advanced technologies and Alberta is fortunate to have a significant number of these individuals but some of them have already been lured away.

ALBERTA AND THE WORLD MARKET

The emergence of computer industry standards several years ago proved to a mixed blessing for much of the Alberta NIC computing community which is largely composed of seismic data processing companies. Although most people saw the merit in moving to a standardized operating system that would not lock them into any one vendors' hardware, they were faced with the prospect of having to do one last major software rewrite. In addition, once their software was ready, they would have to decide which of the many hardware platforms would best suit their needs and capital budget. It is worth noting that of all the new seismic processing systems that are reported to be under development anywhere in the world, there are none that do not use the UNIX operating system as a base.

This period of realization and transition had just begun when the oil price shock of 1986 occurred. With diminished capital pools resulting from reduced work loads, many companies abandoned their conversion efforts and chose to compete with their old technology by simply lowering prices. To many clients in the cash strapped oil industry, this approach was attractive in the short term. Other companies continued with their software development projects either in a reduced manner or by selling the technology or even the whole company to foreign interests.

Some new computers of moderate capacity have been introduced into the local marketplace in recent times but relative to the rest of the world the Alberta companies began to lose the significant advantage they had held for many years. In 1988, two Calgary based contractors made an attempt to reverse this trend by installing supercomputers. Each company acquired an ETA-10 machine from Control Data Corp. but in 1989, after expending considerable money and effort, both organizations were impacted by the collapse of CDC's supercomputer division. This was the same event that prevented the Cyber 205 project at the University of Calgary from reaching its projected potential. The two contractors returned the CDC machines and each company then installed a different brand of mini-supercomputer. These two companies continue to be technology leaders in Alberta with records of successful diversification and each enjoys a moderate amount of international success.

In the world marketplace, prices for seismic processing continue to soften on a unit basis but the number of units and the computational requirements for each unit are increasing dramatically. Alberta companies are declining to bid on some large international contracts because they cannot meet the demands of these big jobs with only their in-house computing resources, even though their systems are relatively idle. If these contractors had access to a supercomputer facility where they could purchase "project specific" computing cycles to meet their overflow requirements, they could increase their export market. There are no significant companies who still generate their own electrical power to carry on their business. They all buy this commodity from the utility company on a demand basis. Many Alberta companies in the NIC computing industry would discover that buying computer cycles as a commodity could give them a significant advantage; particularly in the international markets.

OTHER FACTORS

The Province of Alberta is not lacking in facilities to do "commercial data processing". What we need is a true "supercomputing" service that will provide the needs of a community that must rely on high technology for future prosperity. It is not sufficient to attempt to divert excess commercial cycles to serve the scientific users. As discussed previously, the two kinds of computing are still sufficiently different that any such program would not only achieve mediocre results on a short term basis. Eventually, combining the two types of computing may make sense but at the present time the technology to do this in a rational, production manner, acceptable to the businessman and scientist alike, does not exist. This statement is made based on extensive market research done by an Alberta company and is contrary to representations that have been made to the Task Force by other parties.

In addition to revitalizing the existing data processing business, the availability of cost effective, supercomputing power without the associated high capital cost to get started will encourage the growth of new ventures. Ideas can be developed and programs can be written on inexpensive workstation computers but fully testing a potentially commercial concept often requires large scale facilities. Technology like this already exists in the Alberta business community and it is only a matter of time before the demand for computing becomes evident.

In a project that had its roots in the Cyber 205 project at the University of Calgary, Pulsearch Consolidated Technology Ltd. has almost perfected a software system that will enable them to perform precise structural analysis of oil and gas pipelines. Massive quantities of data acquired by their very successful pipeline "pig" device will be subjected to detailed finite element analysis to determine with a great degree of accuracy if any portion of a pipeline is in danger of rupturing. Even though this technology is still not completed to production standards, the concept is so powerful that its mandatory use has already been considered by some regulatory bodies. Industry has embraced the Pulsearch technology and work has been performed from Thailand to England. In a recent development, they have been contracted to monitor the Aleyeska pipeline across Alaska for the next three years.

A large number of companies operating in Alberta, including many which received Government funding for various purposes, find it necessary to go outside the Province and usually Canada to get their research work done. The lack of supercomputing facilities is often cited as a major reason for not doing this work in Alberta.

Canada and specifically Alberta could choose an easy solution to the question of computing cycles. With reference to the analogy used above, we might decide to remain a "battery driven" economy. Alternatively, we could opt for buying our computer cycles over networks from facilities in the U.S.A. If we select this option then perhaps we should also review our policies on other utility type issues such as telecommunications and broadcasting.

Canada is a sovereign nation and should remain that way. Supercomputing is a strategic requirement for the future prosperity of all regions of Canada. Alberta, with its strong technology base, should lead the way in this vital initiative. The scientific segment of industry forms a strong component of this technology base in Alberta and is responsible for a large part of the progress and diversification achieved to date. They will continue to play a vital role in technological evolution. Alberta based technology that has evolved from the computing industry is recognized and highly regarded worldwide. If we can maintain our talented, innovative work force and provide them with tools like the supercomputer, we will continue to be competitive in the world markets. Supercomputing is not just a matter of academic research in the universities but one of international competitiveness in the new global economy.

A supercomputer installation has great potential to benefit groups like the seismic processing contractors but it would not subsidize them. The present economics of supercomputing are such that each time the price of equipment doubles, the performance increases by a factor of at least four. When this is combined with a high degree of protection from equipment obsolescence, reduced overhead for maintenance and power consumption plus a myriad of other factors it is very likely that many companies in this industrial segment would become clients of a supercomputer facility. Problems will exist but the solutions will be found as a result of the competitive pressures that have always motivated this dynamic group of professionals.

Dr. Fraser Mustard, President of the Canadian Institute for Advanced Research, in a recent meeting with several members of the Task Force on High Performance Computing stated that "--- Canada must stop being a resource based economy and learn to generate wealth from ideas ---". Ideas can be fragile things unless they are nurtured in the proper environment. In the high technology world in which we live today, it is becoming increasingly likely that an idea will have a better chance of maturing to wealth generating status with the aid of supercomputing facilities.

THE UNIVERSITY OF CALGARY'S CYBER 205

With the establishment of the Supercomputing Centre at the University of Calgary, the Province of Alberta, and the University made a bold commitment to one of the major driving forces impacting science and technology. So powerful are these machines, that the computational science and simulation methods employed on them have come to be recognized as the third major avenue of scientific investigation standing equal to traditional theoretical analysis and laboratory experimentation.

The experience at SuperComputing Services (SCS) has been a positive one, in a number of respects, for augmenting the technology infrastructure within Alberta. In particular, hundreds of people have been exposed to advanced supercomputing methods and numerous academic research efforts have been undertaken.

A marketing review of SCS, at this juncture, is timely as questions have arisen regarding the fundamental marketing assumptions that were in place at the time SCS was launched. In particular, concerns regarding the fitness of SCS to succeed as a business focussed on the sale of supercomputer time have been voiced. The dynamics of the supercomputer time-share market are such that this particular concern is of significance. There is widespread excess supply of supercomputer time-share services with demand concentrated in a few key markets and applications. This has produced a slack opportunistic marketplace where price cutting is the norm.

As a university-based foray into the realm of supercomputing, SCS is doing no better than, and no worse than, other university-based supercomputing centres established throughout North America. Through the last decade, a number of universities established supercomputing initiatives on the very simplistic premise that industry would naturally flock to their supercomputer centres and pay rates that guaranteed all costs would be recovered. Without marketing focus, this hoped-for experience never came to pass for most of these centres. The few that have generated a positive cash flow have typically done so by securing long-term contracts with large government agencies and industrial concerns which booked a major portion of the machine resources throughout the contract term.

None of the several supercomputer centres operating today in the U.S.A. are self-supporting, even those with millions of dollars worth of industrial donations and contracts annually. They are all dependent on continuing federal and state funding and for user subsidies for network access.

The only way to treat high performance computers conceptually, is that they are a new and essential part of the infrastructure of any region which claims to be world competitive economically.

The SCS has a number of successes to its credit with respect to the Cyber 205. Among these are:

- usage rose to well over 90% of the total time available after the initial start-up period;
- there were over 200 projects on the Cyber 205 involving 600 users;
- over 190 graduate students were trained in high performance computing;
- there were over 220 paid course attendees from industry and universities;
- over 60 companies made direct use of the Cyber 205, and another 40 made indirect use through consultants; and,
- first-time users were supported in:

reservoir engineering	graphics and animation
pipeline engineering	plant biology and genetics
algorithm development	ice rubble analysis, etc.

Incidentally, both Nova and NovaHusky are still using the system.

Some of the difficulties with the Cyber 205 project were:

- it came on-stream concurrently with a major downturn in the Alberta oil and gas industry;
- the vendor company went out of the supercomputer business and failed to meet its contractual obligations with respect to support and upgrades;
- the aversion of industry to utilizing a university managed facility was not sufficiently appreciated;
- software had to be specially developed to use the machine because the Cyber 205's operating software was not a standard in the petroleum and seismic industries;
- the operating system environment was difficult to understand;
- the infrastructure of support staff for users was too small;
- there was a lack of supplementary hardware, and especially of computers to feed the Cyber 205;
- although the use of the computer to industry was "free", it was not appreciated how much time and effort (and hence money) had to be spent by users before they could utilize the machine;
- there was a limited communications link to the University of Alberta;
- the system was over-scrutinized by Government;
- users were not convinced that the manager (i.e., University of Calgary) was committed to ongoing support of the Centre; and,
- there was insufficient marketing effort by the Centre.

THE UNITED STATES EXPERIENCE[1]

In the U.S. view, the general objectives for access to high performance computers are:

- to provide funds for acquiring computer software and hardware;
- to assist in meeting operational expenses to maintain and manage facilities; and,
- to ensure that scarce computation resources are distributed fairly to the widest range of users.

SUPPORT STRATEGIES

There are four different strategies which the U.S. Government could use to support high performance computers.

1. Fully Government Owned and Operated - apart from high performance computers owned by the Department of Defence, which are generally unavailable to industry, a number of U.S. Government departments and laboratories own high performance computers which they make available for a fee to selected industrial, institutional and academic users. The U.S. Government has not access to date created a facility at their own cost to lease out 100% to non-Government users.

2. Fully or Partially Government Supported Consortia or Institutions - this type of centre, run by either the academic or institutional sector is the most widely used strategy and is much more readily available to non-Government users than are type 1. centres. On average, about 35% of the total capital and operating costs are covered by other than the Federal Government.

3. Government Funding of Individual Supercomputer Projects and Investigators - in this type of centre, the Government supplies funds to users - it is an "indirect" version of 2., which could lead to more competition between centres thereby assisting them to improve their services. However, this version makes it more difficult for centres to pay for upgrading their hardware and software, long-range planning is more difficult than for 2. and the grants to the users may be diverted to buying mini-supercomputers or to graduate student training. About five centres are funded this way, but the strategy is being phased out.

4. Provide Incentives to Private Institutions to Supply High Performance Computer Services - in this concept, the Federal Government would match funds provided by State governments and universities to create High Performance Computer centres. Assistance would also have to be provided for upgrading. There are no existing examples of this strategy.

EXPANDING AND IMPROVING USAGE

The U.S. considers that to advance the science of high performance computing, Federal programs must support five basic objectives.

1. Expand the Capabilities of Human Resources - people capable of using high performance computers are in high demand by industry, businesses and institutions. There is a shortage of people with such skills and any initiative must ensure that the pipeline for delivering trained personnel remains full.

2. Develop Software and Hardware Resources and Technologies - the research and development on the "grand challenges" must continue. Special efforts are needed to ensure progress in developing software to harness the power of high performance computing.

3. Strengthen the Scientific Underpinnings of Computation - support intellectual curiosity directed to the further development of high performance computers.

4. Construct a Broadly Accessible, High Speed Advanced Broadband Network - provide the scientific and educational community access to the facilities, the data and the software needed to explore new applications.

5. Develop New Algorithms for Computational Science - (algorithms are the mathematical formulas used to instruct computers) - new and better algorithms are needed to improve the performance of hardware and software.

DIFFICULTIES AND BARRIERS

Diversity of sources -

High performance computers are expensive to buy and operate. Usage crosses many disciplines. Users are associated with diverse organizations - universities, colleges, government laboratories, industry, etc.

No natural limits -

the potential use for high performance computing capacity appears to be limitless - some problems can completely saturate significant computer resources - it is very difficult to assess needs and to justify expenditures, or to predict future needs.

Disincentives to investment -

recapturing capital and operating costs through fees has not worked because:

- there is a multi-million dollar investment;
- it is difficult to estimate demand for the facility;
- of the question of whether research grants to users will cover their costs; and,
- will researchers use the facility or some other one? (i.e., through network access to other high performance computers).

Note that facility owners charge on the amount of time their facility is used - hence, light usage equals high charges; heavy usage equals low charges (up to saturation of the machine).

EXAMPLES OF STATES SUPPORTING SUPERCOMPUTER CENTRES

(Note: This list is incomplete)

North Carolina	New York	Illinois
Pennsylvania	California	Indiana
Minnesota	Colorado	Ohio
Texas	Alabama	

(1) The material in this section was adapted from "Seeking Solutions: High Performance Computing for Science". U.S. Congress, Office of Technology Assessment, April 1991.

CANADIAN SUPERCOMPUTER INSTALLATIONS[1]

ATMOSPHERIC ENVIRONMENT SERVICE (DORVAL)

HYDRO QUEBEC

ONTARIO CENTRE FOR LARGE SCALE COMPUTING (OCLSC), UNIVERSITY OF TORONTO

SHELL CANADA RESOURCES (CALGARY)

A "CLASSIFIED" INSTALLATION

Note: All installations, as at January 1, 1991, involved Cray machines, with the exception of the obsolete Cyber 205 at the University of Calgary. The Dorval machine is in the process of being replaced by an NEC machine. The Toronto machine is obsolescent and a proposal is being prepared which will suggest that the Province of Ontario assist to the extent of several tens of millions of dollars in upgrading OCLSC.

(1) Material supplied by Fujitsu Canada Inc., the Atmospheric Environment Service and OCLSC.

BIBLIOGRAPHY

1. Brooking, James, "Statewide Supercomputing in North Carolina: From Zero to a Gigaflop in Two Days", North Carolina Supercomputing Centre, Supercomputing Symposium '91.

2. Commission of the European Communities - Report of the EEC Working Group on High Performance Computing.

3. Congress of the United States Office of Technology Assessment - "Seeking Solutions: High Performance Computing for Science", April 1991.

4. Fisher, Joseph and Ramakrishna Rao, "Instruction-Level Parallel Processing", Science, Vol. 253, September 13, 1991, pp. 1233-1241.

5. Freedman, David, "A Painless Route to Parallel Computing", Science, Vol. 253, September 13, 1991, pp. 1209-10.

6. George, Alan, Andrew Bjerring, and Pardeep Ahluwalia, "National Initiative for Scientific Computing: A Proposal for Implementation", a submission to NSERC, June 1991.

7. George, Alan, et al, Report of the ad hoc Committee on Research Computation to NSERC, September 1990.

8. Ghent-Mallett, Jocelyn, Presentation on ISTCNet to Alberta officials, June 1991.

9. Gore, Al (Senator), A Bill to Provide for a Coordinated Federal Research Program to Ensure Continued United States leadership in High Performance Computing, S-272, Calendar No. 87, 102nd Congress, 1st Session, May 16, 1991.

10. Hickling Associates, "Feasibility Study of a National High Speed Communications Network for Research, Development and Education, March 1990.

11. Karp, Alan, "The Global Computer", IBM Scientific Centre, Palo, Alto, Document no. G32-3544, December 18, 1990.

12. Lundstrom, Stephen, "The Future of Supercomputing: Fact and Fantasy", PARSA Technical Report No. 15, Hawaii International Conference on System Sciences-24, January 8, 1991.

13. Management Board of the OCLSC - A Proposal for the Extension of Provincial Support for the Ontario Centre for Large Scale Computation, January 9, 1991.

14. Minnesota Supercomputing Centre, publicity package, Summer 1991.

15. North Carolina Supercomputing Centre - package of publicity material, Summer 1991.

16. Office of Science and Technology Policy - "Grand Challenges: High Performance Computing and Communications." Supplement to the President's Fiscal Year 1992 Budget.

17. Office of Science and Technology Policy - The Federal High Performance Computing Program, Executive Office of the President, September 8, 1989.

18. Office of Science and Technology Policy - A Research and Development Strategy for High Performance Computing, November 1987.

19. Peterson, Victor and James Arnold, "The Impact of Supercomputers on Experimentation: A View from a National Laboratory", 1985 ASEE Annual Conference Proceedings.

20. Pittsburgh Supercomputing Centre - Projects in Scientific Computing, 1989-90, July 1990.

21. Quimby, David, "Business Supercomputing: An Emerging Reality", SRI International, August 1989.

22. Raveche, Harold, Duncan Lawrie, and Alvin Despain, "A National Computing Initiative: The Agenda for Leadership". Report of the Panel on Research Issues in Large-Scale Computational Science and Engineering, Society for Industrial and Applied Mathematics, Philadelphia, 1987.

23. Scientific American - Special issue on communications, computers and networks, September 1991.